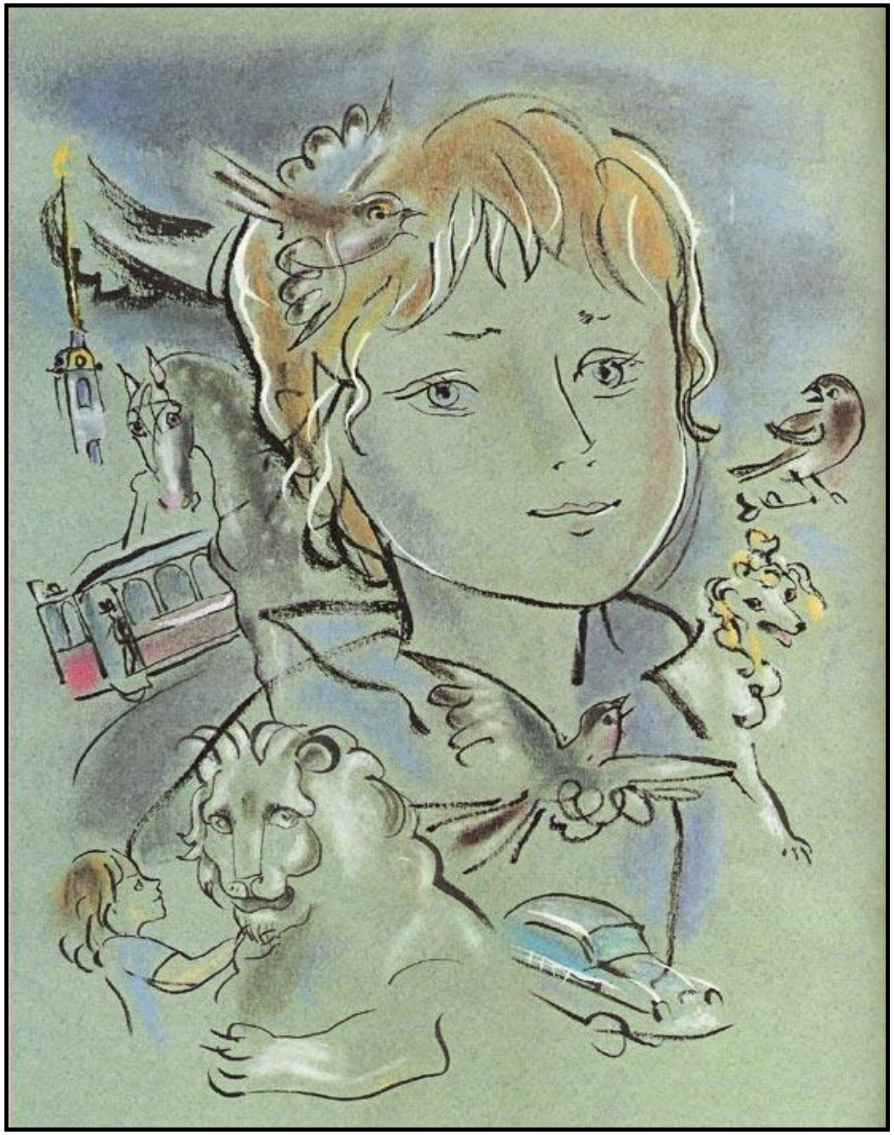
Object Recognition: Conceptual Issues



Slides adapted from Fei-Fei Li, Rob Fergus, Antonio Torralba, and K. Grauman

Issues in recognition

The statistical viewpoint

Generative vs. discriminative methods

Model representation

Supervised vs. unsupervised methods

Different recognition tasks

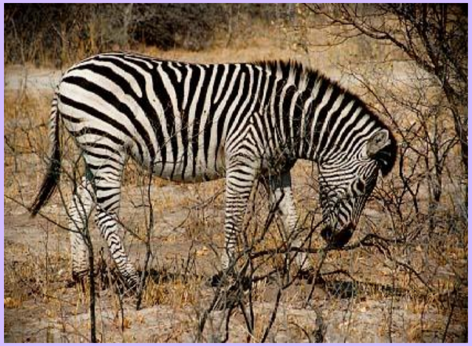
Datasets

Object categorization: the statistical viewpoint

- MAP decision: $p(zebra \mid image)$

vs.

$p(no\ zebra \mid image)$

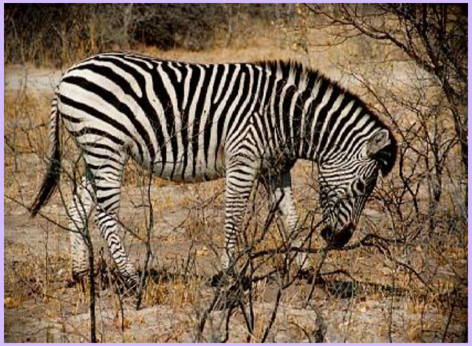


Object categorization: the statistical viewpoint

- MAP decision: $p(zebra \mid image)$

vs.

$$p(no\ zebra \mid image)$$



- Bayes rule:

$$\underbrace{p(zebra \mid image)}_{\text{posterior}} \propto \underbrace{p(image \mid zebra)}_{\text{likelihood}} \underbrace{p(zebra)}_{\text{prior}}$$

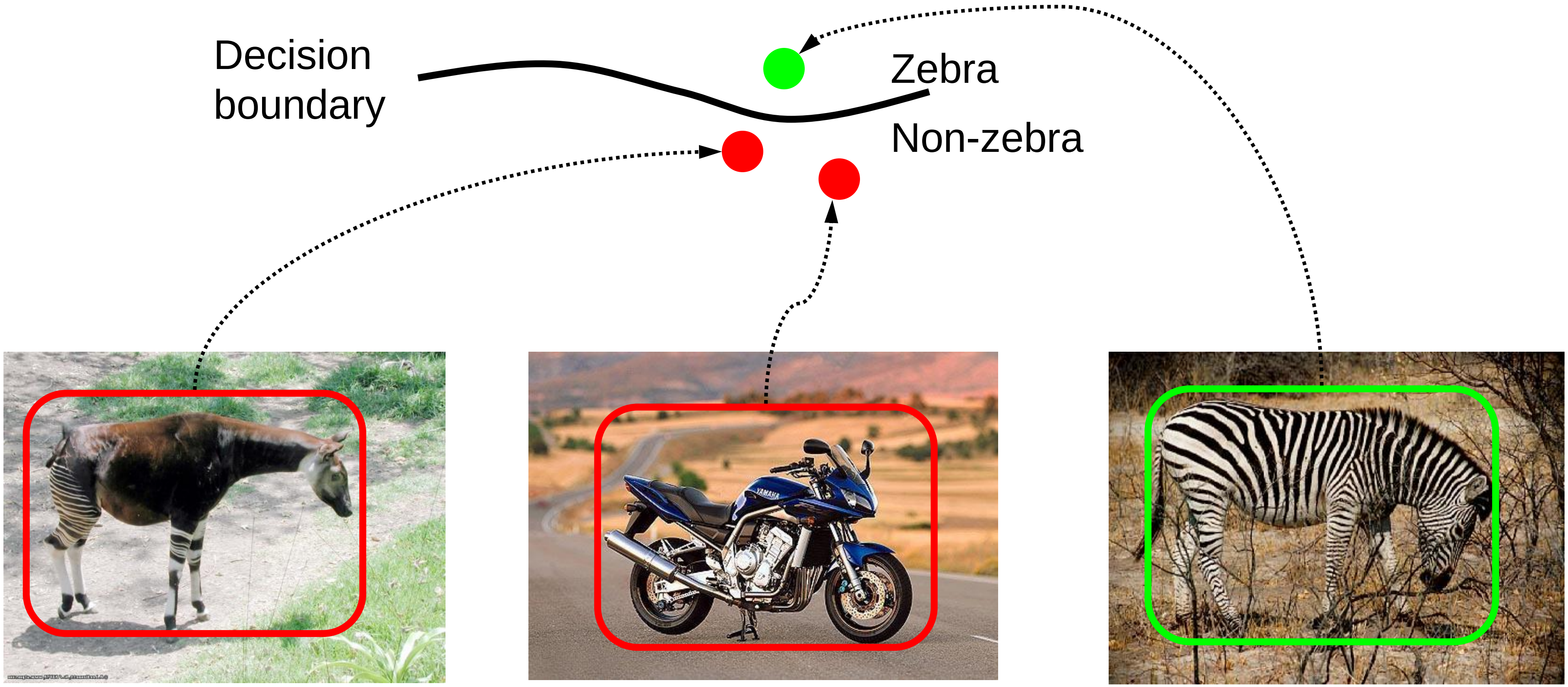
Object categorization: the statistical viewpoint

$$\underbrace{p(zebra \mid image)}_{\text{posterior}} \propto \underbrace{p(image \mid zebra)}_{\text{likelihood}}\,\underbrace{p(zebra)}_{\text{prior}}$$

- **Discriminative methods: model posterior**
- **Generative methods: model likelihood and prior**

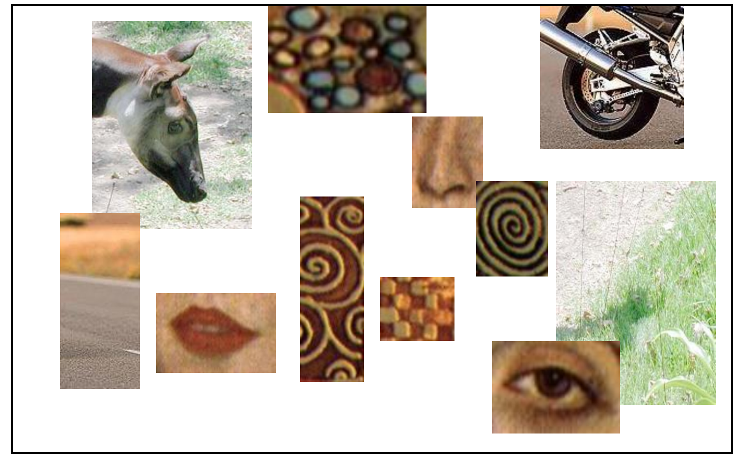
Discriminative methods

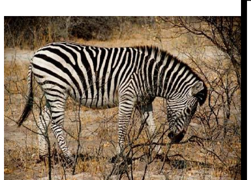
- Direct modeling of $p(zebra \mid image)$



Generative methods

- Model $p(image \mid zebra)$ and $p(image \mid no\ zebra)$



	$p(image \mid zebra)$	$p(image \mid no\ zebra)$
	Low	Middle
	High	Middle→Low

Generative vs. discriminative methods

- Generative methods
  + Interpretable
  + Can be learned using images from just a single category
  – Sometimes we don’t need to model the likelihood when all we want is to make a decision

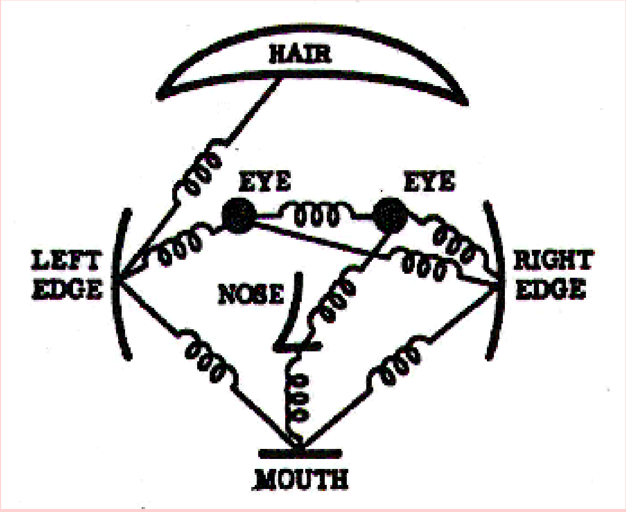
- Discriminative methods
  + Efficient
  + Often produce better classification rates
  – Can be hard to interpret
  – Require positive and negative training data

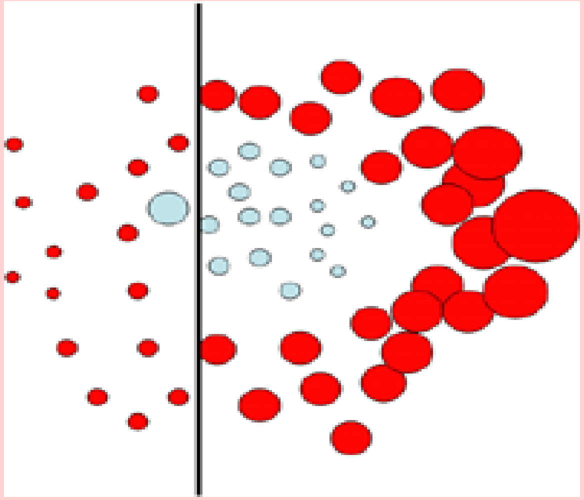
Issues for statistical recognition

- Representation
  - How to model an object category
- Learning
  - How to find the parameters of the model, given training data
- Recognition
  - How the model is to be used on novel data

Representation

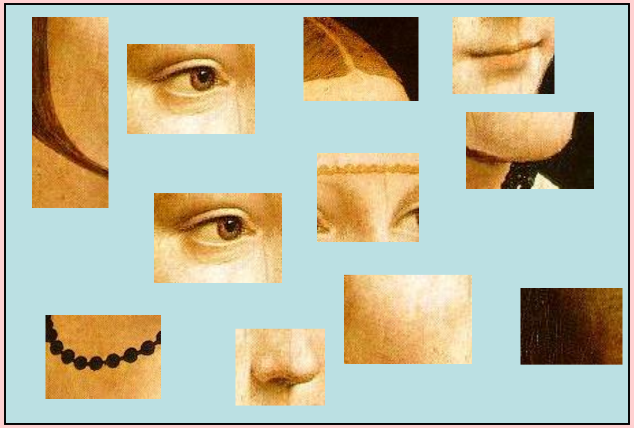
– Generative / discriminative / hybrid

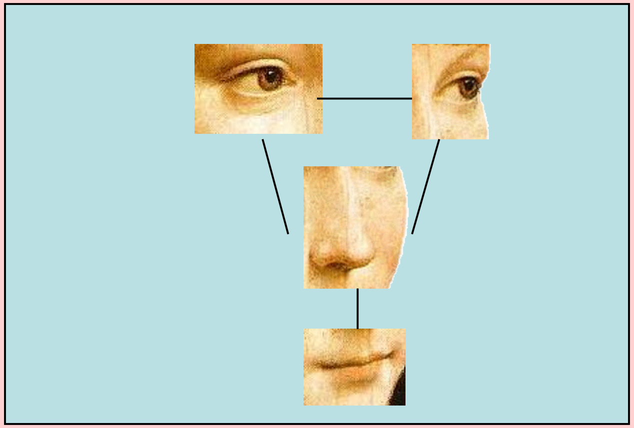




Representation

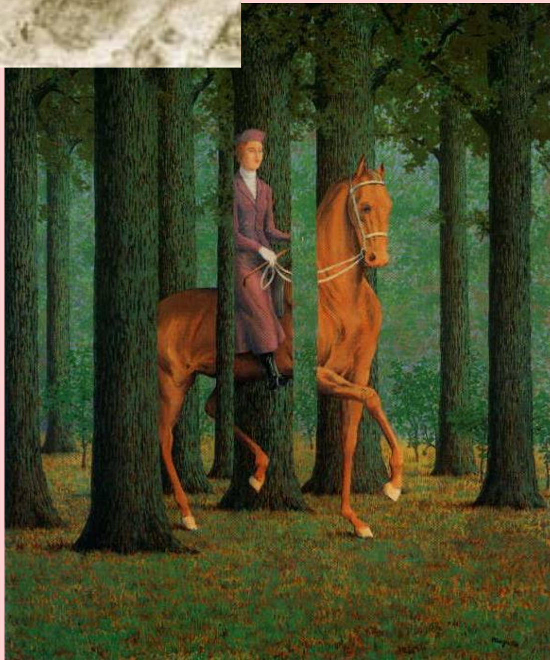
– Generative / discriminative / hybrid
– Appearance only or location and appearance





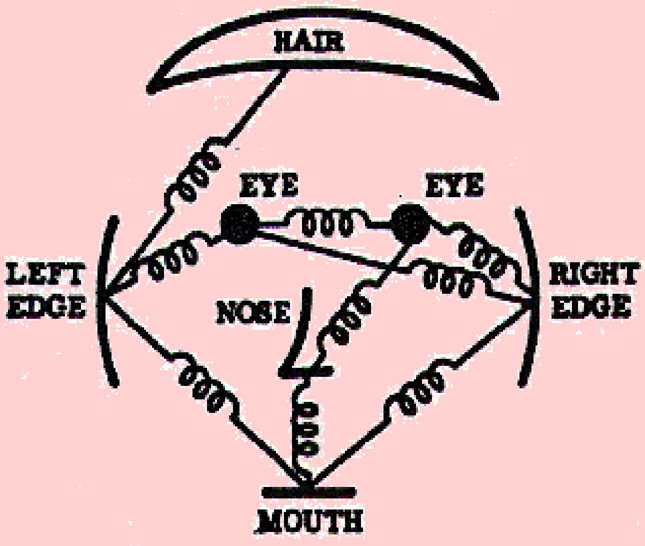
Representation

- Generative / discriminative / hybrid
- Appearance only or location and appearance
- Invariances
  - View point
  - Illumination
  - Occlusion
  - Scale
  - Deformation
  - Clutter
  - etc.



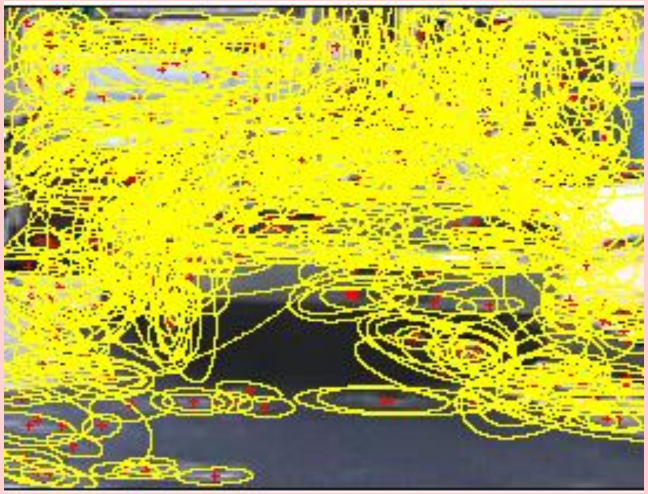
Representation

- – Generative / discriminative / hybrid
- – Appearance only or location and appearance
- – Invariances
- – Part-based, global, sliding window



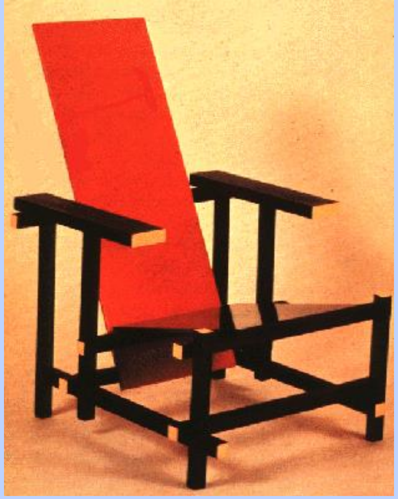
Representation

- Generative / discriminative / hybrid
- Appearance only or location and appearance
- Invariances
- Part-based, global, sliding window
- Use set of features or each pixel in image



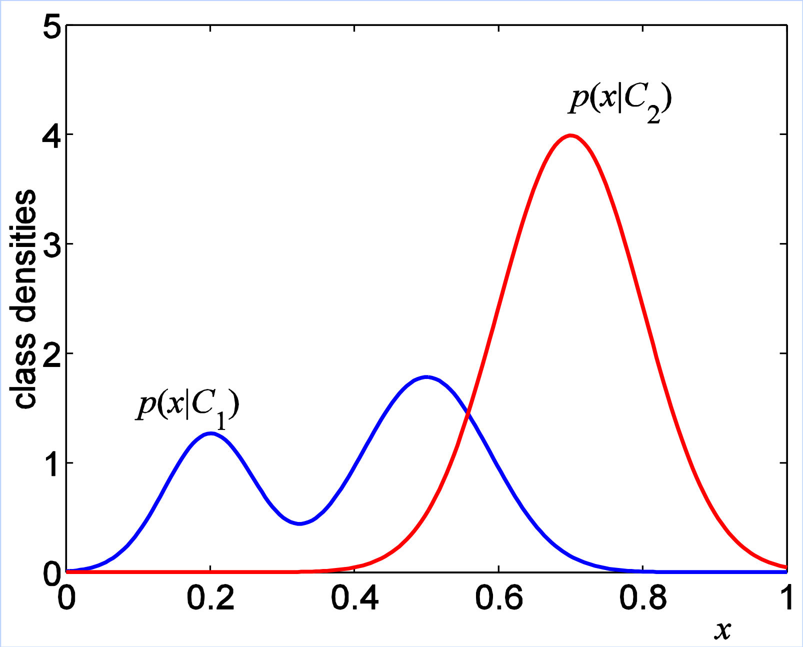
Learning

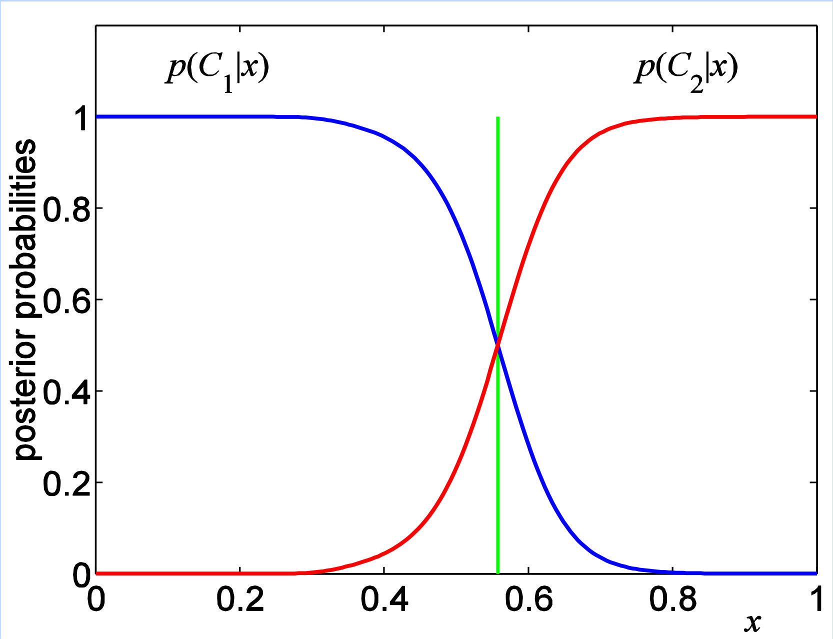
- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning



Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- Methods of training: generative vs. discriminative





Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- Methods of training: generative vs. discriminative
- Generalization, overfitting, bias vs. variance

Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- Methods of training: generative vs. discriminative
- Generalization, overfitting, bias vs. variance
- Level of supervision
  - Manual segmentation; bounding box; image labels; noisy labels
  - Task-dependent

Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- Methods of training: generative vs. discriminative
- Generalization, overfitting, bias vs. variance
- Level of supervision
  - Manual segmentation; bounding box; image labels; noisy labels
  - Task-dependent

Contains a motorbike

Learning

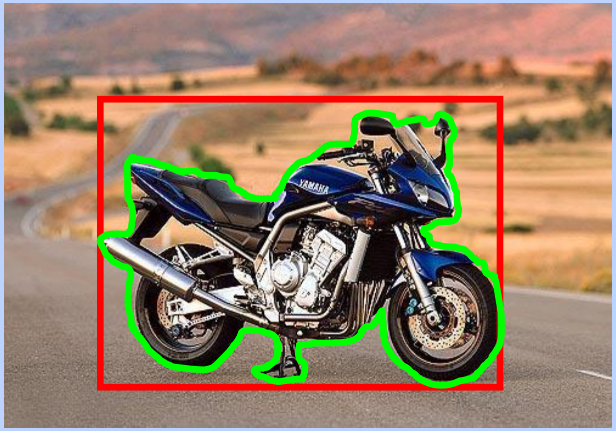
- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- Methods of training: generative vs. discriminative
- Generalization, overfitting, bias vs. variance
- Level of supervision
  - Manual segmentation; bounding box; image labels; noisy labels
  - Task-dependent

Contains a motorbike

Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- Methods of training: generative vs. discriminative
- Generalization, overfitting, bias vs. variance
- Level of supervision
  - Manual segmentation; bounding box; image labels; noisy labels
  - Task-dependent

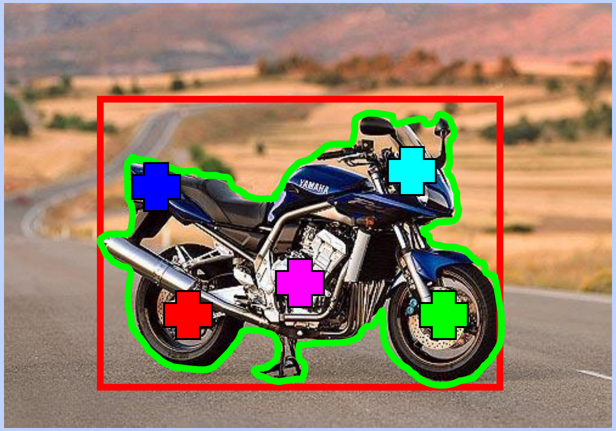
Contains a motorbike



Learning

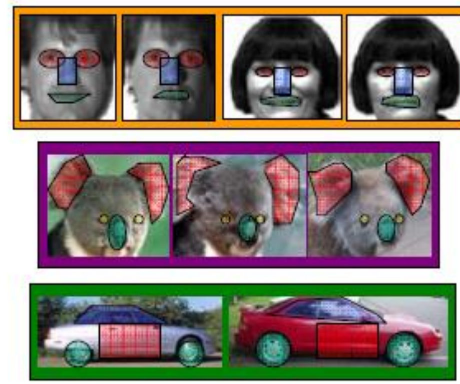
- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- Methods of training: generative vs. discriminative
- Generalization, overfitting, bias vs. variance
- Level of supervision
  - Manual segmentation; bounding box; image labels; noisy labels
  - Task-dependent

Contains a motorbike



Spectrum of supervision

Less **More**



Unsupervised

"Weakly" supervised

Supervised

Definition depends on task

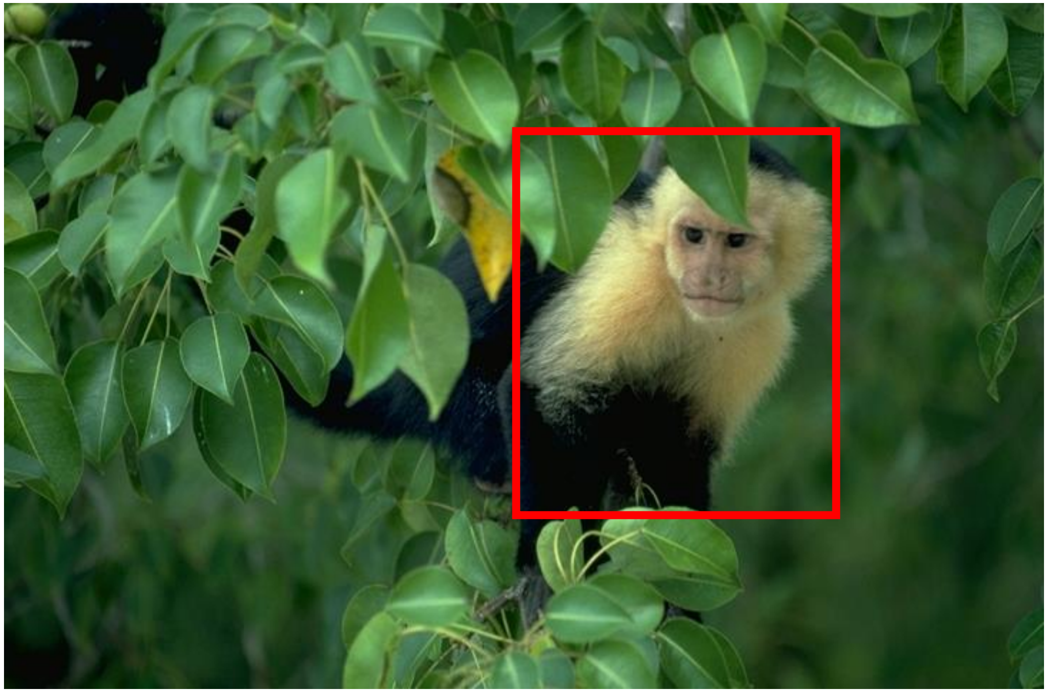
What task?

- Classification
  - Object present/absent in image
  - Background may be correlated with object

What task?

- Classification
  - Object present/absent in image
  - Background may be correlated with object

- Localization / Detection
  - Localize object within the frame
  - Bounding box or pixel-level segmentation



Datasets

- Circa 2001: 5 categories, 100s of images per category
- Circa 2004: 101 categories
- Today: thousands of categories, tens of thousands of images

Caltech 101 & 256

Fei-Fei, Fergus, Perona, 2004

Griffin, Holub, Perona, 2007

Collecting datasets (towards 10^{6-7} examples)

- ESP game (CMU)

Luis Von Ahn and Laura Dabbish 2004

- LabelMe (MIT)

Russell, Torralba, Freeman, 2005

- StreetScenes (CBCL-MIT)

Bileschi, Poggio, 2006

- WhatWhere (Caltech)

Perona et al, 2007

- PASCAL challenge

2006, 2007

- Lotus Hill Institute

Song-Chun Zhu et al 2007

- Tiny Images (MIT)

Torralba, Fergus & Freeman 2007

Labeling with games

Figure 1. Partners agreeing on an image in the ESP Game. *Neither player can see the other's guesses.*

Figure 2. Peekaboom. *"Peek" tries to guess the word associated with an image slowly revealed by "Boom."*

L. von Ahn, L. Dabbish, 2004; L. von Ahn, R. Liu and M. Blum, 2006

Lotus Hill Research Institute image corpus

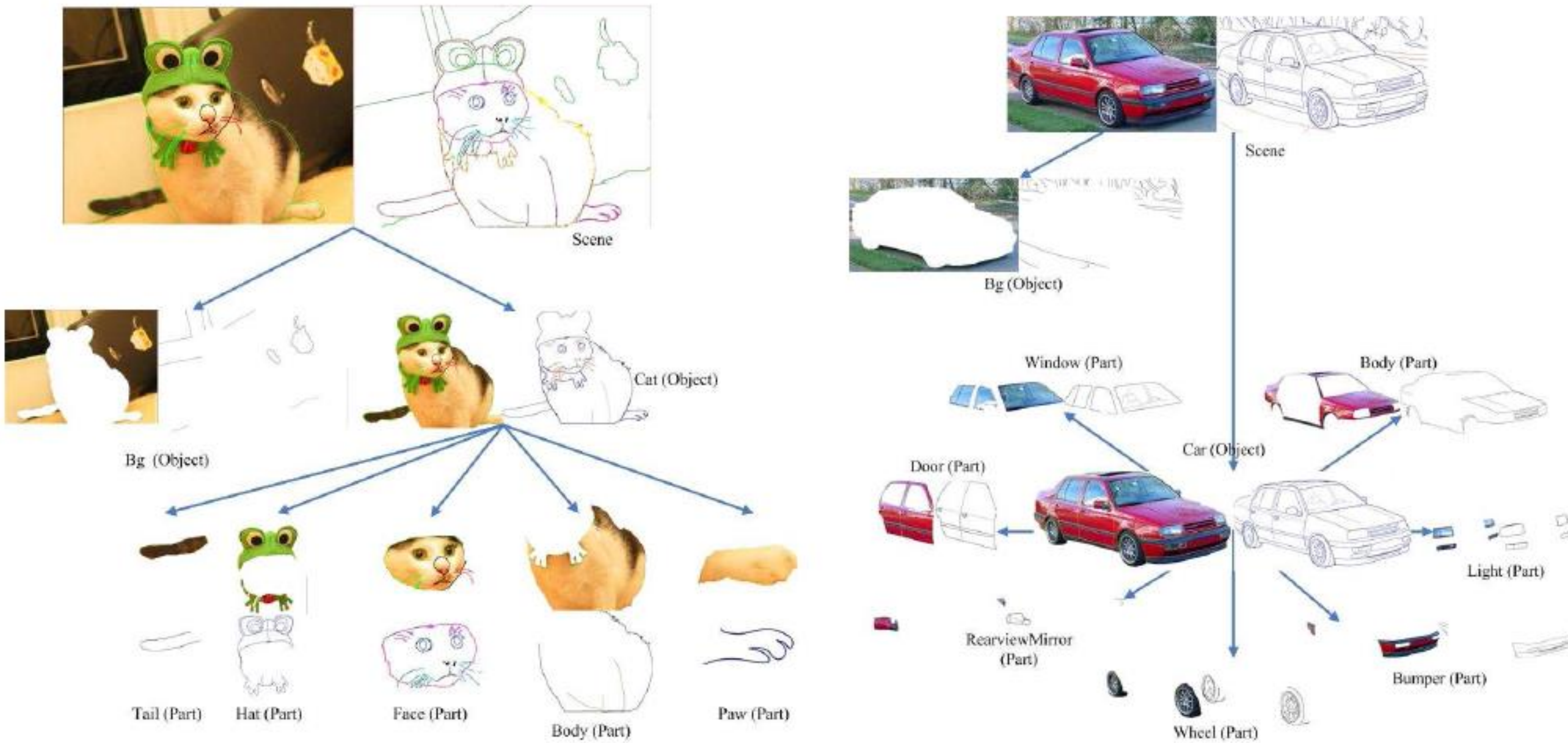


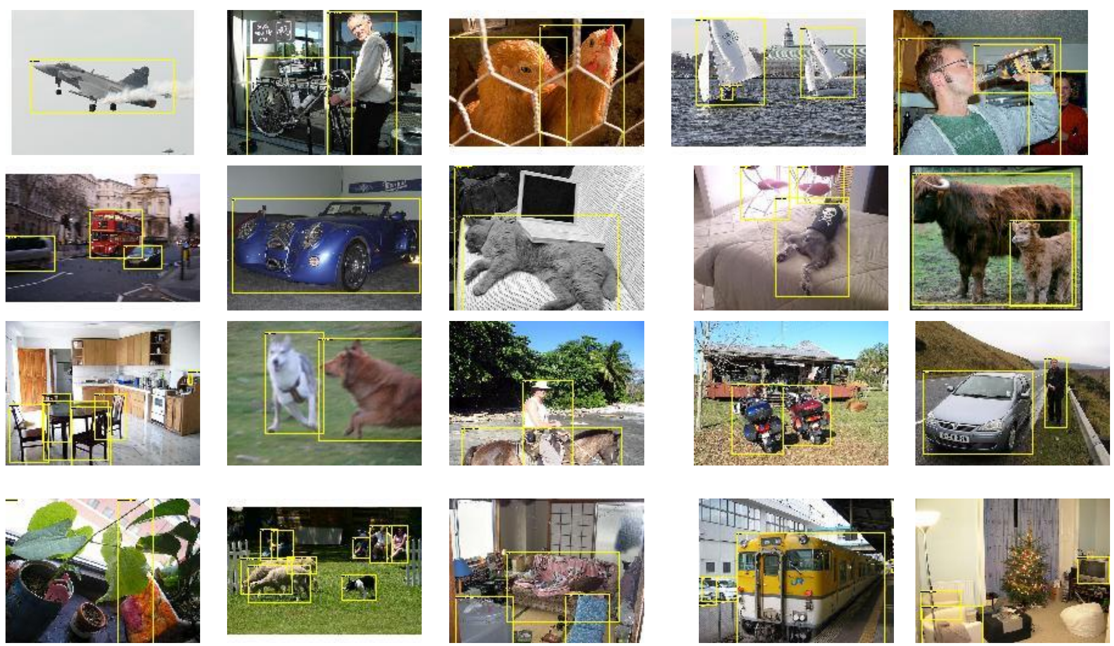
Figure 5: Two examples of the parse trees (cat and car) in the Lotus Hill Research Institute image corpus. From [87].

Z.Y. Yao, X. Yang, and S.C. Zhu, 2007

The PASCAL Visual Object Classes Challenge 2007

The twenty object classes that have been selected are:

Person: person
Animal: bird, cat, cow, dog, horse, sheep
Vehicle: aeroplane, bicycle, boat, bus, car, motorbike, train
Indoor: bottle, chair, dining table, potted plant, sofa, tv/monitor



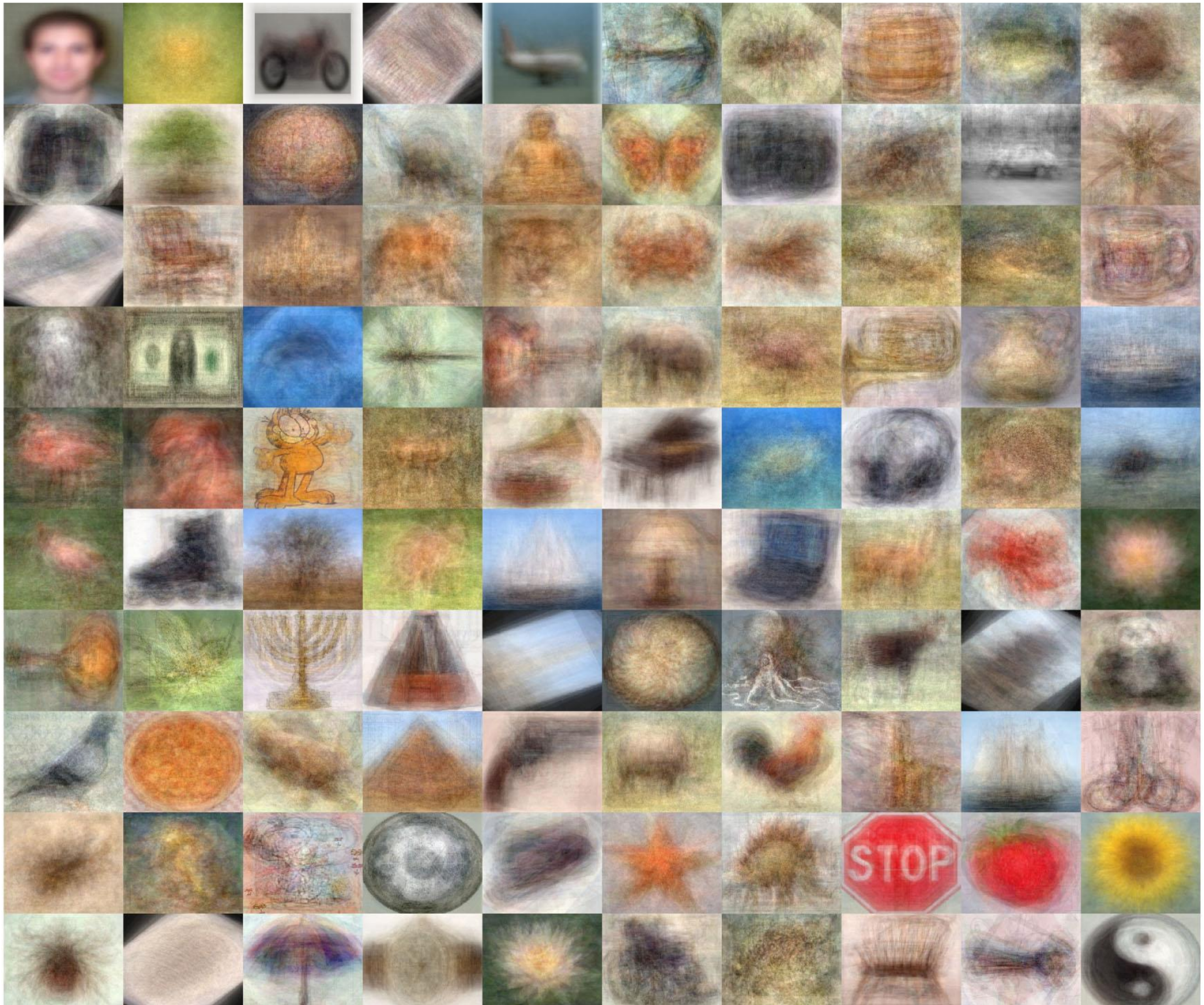
LabelMe

Russell, Torralba, Freman, 2005

Dataset issues

- How large is the degree of intra-class variability?
- How “confusable” are the classes?
- Is there bias introduced by the background? I.e., can we “cheat” just by looking at the background and not the object?

Caltech-101



Summary

- Recognition is the “grand challenge” of computer vision
- History
  - Geometric methods
  - Appearance-based methods
  - Sliding window approaches
  - Local features
  - Parts-and-shape approaches
  - Bag-of-features approaches
- Issues
  - Generative vs. discriminative models
  - Supervised vs. unsupervised methods
  - Tasks, datasets